Best practice for single-trial detection of event-related potentials:Application to brain-computer interfaces

International Journal of Psychophysiology
journal homepage: www.elsevier.com

Best practice for single-trial detection of event-related potentials: Application to brain-computer interfaces

Hubert Cecotti[a,**], Anthony J. Ries[b]

[a] *Faculty of Computing and Engineering, Ulster University, Magee campus, Londonderry, BT48 7JL, Northern Ireland, UK.*
[b] *Human Research and Engineering Directorate, US Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA.*

ARTICLE INFO

Article history:
Received 1 May 201X
Received in final form 10 May 201X
Accepted 13 May 201X
Available online 15 May 201X

Communicated by S. Abcd

Event-related potentials
Brain-Computer Interface
Biomedical engineering
Spatial filtering
Multivariate pattern analysis
Classification

ABSTRACT

The detection of event-related potentials (ERPs) in the electroencephalogram (EEG) signal is a fundamental component in non-invasive brain-computer interface (BCI) research, and in modern cognitive neuroscience studies. Whereas the grand average response across trials provides an estimation of essential characteristics of a brain-evoked response, an estimation of the differences between trials for a particular type of stimulus can provide key insight about the brain dynamics and possible origins of the brain response. The research in ERP single-trial detection has been mainly driven by applications in biomedical engineering, with an interest from machine learning and signal processing groups that test novel methods on noisy signals. Efficient single-trial detection techniques require processing steps that include temporal filtering, spatial filtering, and classification. In this paper, we review the current state-of-the-art methods for single-trial detection of event-related potentials with applications in BCI. Efficient single-trial detection techniques should embed simple yet efficient functions requiring as few hyper-parameters as possible. The focus of this paper is on methods that do not include a large number of hyper-parameters and can be easily implemented with datasets containing a limited number of trials. A benchmark of different classification methods is proposed on a database recorded from sixteen healthy subjects during a rapid serial visual presentation task. The results support the conclusion that single-trial detection can be achieved with an area under the ROC curve superior to 0.9 with less than ten sensors and 20 trials corresponding to the presentation of a target. Whereas the number of sensors is not a key element for efficient single-trial detection, the number of trials must be carefully chosen for creating a robust classifier.

1. Introduction

An event-related potential (ERP) is the measured brain response evoked by specific sensory, cognitive, or motor event. More generally, it is any stereotyped electrophysiological response to a stimulus. The ERP technique provides a powerful non-invasive tool for exploring the human brain, particularly for research related to the temporal measurement of cognitive mechanisms [52]. An ERP component corresponds to the scalp-recorded neural activity generated from cortical sources and can be reliably measured using electroencephalography (EEG), a procedure that measures electrical activity of the brain over time using electrodes placed on the scalp [51]. Since EEG measurements reflect thousands of simultaneous post-synaptic neural activations, the brain response to a single stimulus or event of interest is not usually visible in the EEG recording of a single trial [16]. Therefore, experimenters typically average many trials

**Corresponding author: Tel.: +44-287-167-5276
e-mail: h.cecotti@ulster.ac.uk (Hubert Cecotti)

together in order to see the brain's response to a stimulus. This causes random brain activity to be averaged out, and the relevant waveform to remain, i.e. the ERP. One of the most widely studied ERP components, first reported over 50 years ago, is the P3 ERP component, a large positive wave that peaks around 300 ms after the stimulus onset [77]. The P3 is often used to index processing related to stimulus categorization and as an input signal in many brain computer interface (BCI) systems for both patients and healthy individuals [25, 26]. The P3 has been traditionally studied using a two-stimulus oddball task where an infrequent "oddball" target stimulus is presented among a series of frequent non-target distractor stimuli where only the infrequent target stimulus requires a response by the observer [67]. The targets of interest produce significantly larger P300 amplitudes than non-target distracters and do so even when both targets and non-target distracters are equiprobable [41]. The latency of the P3 is correlated with the reaction time to a target stimulus making the latency at which this component peaks a useful metric for estimating the time it takes to evaluate and categorize a target stimulus [22, 29]. Targets that are easier to categorize produce faster reaction times and earlier P3 peak latencies than more difficult targets. For reviews on the P3 ERP see [54, 67]. Other ERPs components such as the N2 (a negative wave that peaks 150-350ms post-stimulus), which usually precedes the P3, have been used with single-trial detection as a single component, or with the P3. This ERP component has been extensively studied to find out its relationship with selective attention to specify stimulus location in certain area of the visual field [46]. In the subsequent sections of this paper, the ERP definition includes the brain response relative to any sensory, cognitive, or motor event stimulus. Hence, it includes both visual evoked potentials (VEPs), and auditory evoked responses (AEPs) [45]. Finally, to limit the scope of this paper dedicated to ERP single-trial detection, this paper focuses on multivariate pattern analysis (MVPA) methods that allow the extraction of meaningful and reliable information about the presence of a particular ERP at the single-trial level (e.g. P3; target vs. non-target trial). This differs from methods that extract relevant information from a population of trials by using features from each individual trial (e.g. amplitude or latency measurements across trials). In this later case, while the analysis is performed at the single-trial level, only the group analysis can provide information about neural processes.

In typical ERP analysis in cognitive neuroscience, the grand average response across trials is used to analyze and compare differences between ERP characteristics (amplitude, latency) across subjects. Furthermore, multivariate approaches have been used since the early days of ERP analysis (e.g. the quantitative analysis of averaged evoked potentials [23, 24]). In addition, the graphical representation of sorted ERPs based on the latency or the amplitude of a particular component provides a tool to analyze brain evoked response. However, it is important to be cautious about the interpretation of ERP experiments. In most of the ERP experiments, the different ERP waveforms are isolated by using the grand average response. Yet, the variations across trials may not be captured by the grand average response, which may provide a biased view of the single-trial waveforms. This effect is enhanced when ERP component latencies vary significantly across trials. Hence, it is ideally better to not assume that an averaged ERP waveform represents the single-trial waveform prototype. Only a few studies have taken this information into account [58], by extracting shift-invariant features from the EEG signal [9].

Modeling trial-to-trial variability in EEG signal has become a major focus in single-trial classification. Latencies at the single-trial level are typically by peak-picking and template-matching [76]. Already in [47], it was shown at the single-trial level that the latency of P3 corresponds to stimulus evaluation time and is independent of response selection. In [21], they have developed a method based on an ERP-image visualization tool in characteristic such as potential and spectral power are represented as color coded horizontal lines that are then stacked to form a 2-D colored image. Moving-window smoothing across trial epochs can make otherwise hidden ERP features in the data more perceptible. Stacking trials in different orders, for example ordered by subject reaction time, by context-related information such as inter-stimulus interval, or some other characteristic of the data (e.g., latency-window mean power or phase of some EEG source) can reveal aspects of the multifold complexities of trial-to-trial EEG data variability.

In this paper, we define a complete ERP signal-trial detection pipeline that includes temporal filtering, spatial filtering, and classification. For each component, we review the best methods and the best parameters that are currently used. We compare the performance of several state-of-the art techniques from the BCI literature in order to assess the performance of these different approaches to highlight the most efficient pre-processing steps and classification procedure. To compare the different techniques, we consider a database of 16 healthy subjects performing a visual target detection task, and where it can be expected to find a major N2 and P3 component with a high amplitude. Moreover, this task illustrates how oddball paradigms have evolved to complex and more applied problems that can be applied in novel BCI systems for potential threat detection with a direct impact on society. The remainder of this paper is organized as follows: First, we present how ERPs detection is used in BCI in Sections 2 and 3. After the description of the experimental protocol related to the data used in this paper, in Section 4, we describe the system architecture in Section 5. The pre-processing and classification techniques are detailed in Sections 6 and 7. Finally, the results are presented in Section 8 and discussed in Section 9.

2. Brain-computer interface

Brain-Computer Interface (BCI) or Brain-Machine Interface (BMI) systems have been introduced as a new means of communication for severely disabled people who are unable to communicate with conventional devices (e.g. mouse, keyboard, switch), for rehabilitation purposes [60, 82], and to enhance performance of healthy individuals [49]. BCIs based on ERP detection require subjects to pay attention to a specific sequence of stimuli (typically visual or auditory) in order to produce a robust and detectable ERP. The stability of the spatial distribution, the amplitude, and the latency of a brain evoked responses are key features that allow robust single-trial detection. It is now possible

to reliably detect brain evoked responses using efficient signal processing methods that denoise the signal and enhance its main discriminant characteristics. This principle has been used in BCI to detect specific event-related potentials [82]. Virtual keyboards based on the detection of ERPs have been used in BCI, the most famous variation is the P300 speller [27], and new variations based on other ERP components have been proposed [36]. A large number of studies have been dedicated directly to the P300 speller, in order to understand the impact of its parameters, and how this system can be efficiently used [31]. While the P300 ERP is relatively stable in P300 speller paradigms, accurate and reliable detection of the specific neural responses often requires averaging multiple responses. For instance, it is common that about ten trials are averaged in BCI virtual keyboards to optimize the accuracy [15]. The requirement of several trials is mainly due to the noise in the signal coming from eye movements, muscular contractions, and ongoing brain activity that is unrelated to the experimental task. Although averaging the signal from multiple brain responses can increase the efficiency of detection, it also decreases the information transfer rate of the BCI due to the increase of time to acquire additional trials that are needed to reach a robust decision [7]. Moreover, there exist tasks where it is not possible to repeat the visual stimuli: they appear only one time [9]. It happens in paradigms where the repetition may have an effect on the brain evoked response (use of memory), or when the application does not allow the repetition of the stimuli, e.g. when a subject watches a video; each frame of the video is presented only one time. In situations where novel incoming stimuli are presented in real-time, it may not be possible to repeat the presentation of visual stimuli in order to combine the decision scores from their corresponding brain responses. For this reason, single-trial detection has to be used for target detection where it is not possible to determine if an image belongs to a target or a non-target class by considering multiple presentations of the same image. Yet, if images can be presented several times, it is possible to combine the decision outputs from the different presentations like in the P300 speller [27]. Thus, the real time constraint justifies the necessity to find new strategies for increasing the performance of single-trial detection. Finally, the N2pc (posterior-contralateral) has been used in some recent BCI studies [2, 59, 75] to extract information about the spatial location of targets in images: a stronger deflection amplitude is expected in the area of the visual cortex which is opposite to the location of the visual stimulus.

3. Rapid Serial Visual Presentation tasks

Rapid Serial Visual Presentation (RSVP) tasks are well suited for ERP based BCI paradigms [37, 73]. In the RSVP paradigm, a rapid sequence of images is presented to subjects in the same location on a screen, which makes this type of BCI gaze-independent [17, 34, 68]. The presentation rate of the RSVP sequence is determined by the stimulus onset asynchrony (SOA) (in ms) or the frequency of image presentation (in Hz) (see Fig. 1). The stream of images contains different types of visual stimuli, which can be clustered into different groups. With two groups of stimuli, there is typically a set of images representing

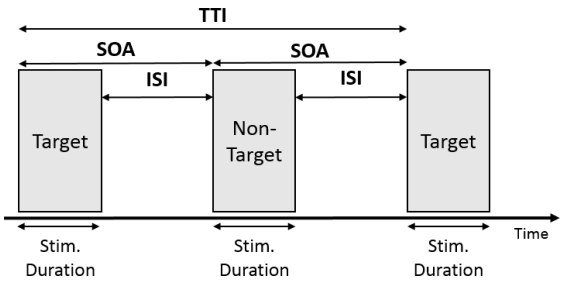


Fig. 1. Main parameters during the presentation of the stimuli during an RSVP task. ISI: inter-stimulus interval, SOA: stimulus onset asynchrony, TTI: target-to-target interval.

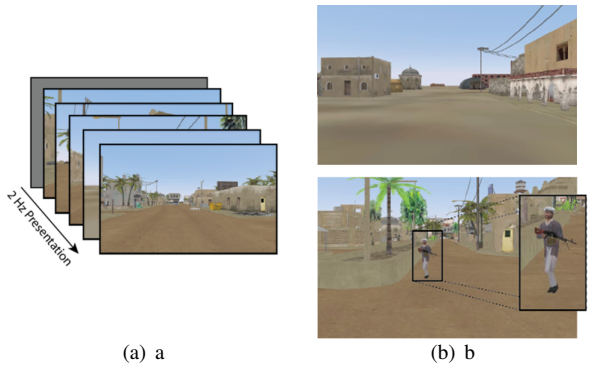
targets (images that are relevant to the task and for the subject), and a set of images representing non-targets (images that the subject should ignore or that are irrelevant). The number of target images is significantly lower than the number of non-target images. This difference of distribution between target and non-target images is needed to obtain a high target-to-target interval (TTI). With more non-target images and a random normal distribution of images over time, the TTI will increase, leading to an increase of the P3 amplitude, and an increase in P3 amplitude would likely result in better discriminant features for the classifier. Previous studies have shown that visual processing needed during a go/no-go categorization task can be achieved under 150 ms [78]. In [19], it was suggested that the inter-stimulus interval (ISI) and probability do not independently affect P3 amplitude, and that TTI offers a strong explanation of the reported relations between P3 amplitude and both ISI and probability. Moreover, variations of the SOA have several important consequences on the ERP: reducing the SOA increases the perceptual difficulty of the task, increasing the likelihood for errors [44, 69], decreasing the SOA will involve a decrease of the TTI. In BCI, RSVP tasks have been used to increase information throughput for image analysis by sorting and triaging images based on the characteristics of the evoked responses [3, 30, 64–66, 83], and face recognition tasks [6, 79].

For the creation of a single-trial method, it is important to take into account the aforementioned concerns. First, the distribution of classes is typically unbalanced for binary classification as one class will have a larger number of trials than the other one, e.g. non-target vs. target in oddball paradigm where the target has a low probability (e.g. 10%). Second, the number of available trials for training a classifier, or tuning parameters of the model through grid search is limited. Because the EEG signal is noisy and the number of available trials for creating a model is typically low, feature reduction and selection methods must be applied to train classifiers. The next section presents the RSVP task that is considered to illustrate the use of ERP single-trial classification.

4. Experimental protocol

4.1. Subjects

Eighteen participants volunteered for the study. All participants provided written informed consent, reported normal or



(a) a (b) b

Fig. 2. (a) Rapid Serial Visual Presentation task. (b) Representative examples of stimuli on target (bottom) and non-target trials (top). The inset with a target is depicted for illustration purposes, and did not appear in actual stimuli.

corrected-to-normal vision, and reported no history of neurological problems. Two participants were excluded from analysis because there were excessive noise artifacts in the recorded EEG signal. The resulting 16 participants had an average age of 33.5 years (13 males, 15 right handed). The voluntary, fully informed consent of the persons used in this research was obtained as required by federal and Army regulations. The investigator has adhered to Army policies for the protection of human subjects [80, 81].

4.2. Visual stimuli and procedure

The data were previously used in [12, 57], and are briefly described here. Participants were seated 75 cm from a Dell P2210 monitor, and viewed a series of simulated images from a desert metropolitan environment in a rapid serial visual presentation (RSVP) paradigm (Fig. 2(a)). Images (960×600 pixels, 96 dpi, subtending 36.3×22.5) were presented using E-prime software on a Dell Precision T7400 PC. Images were presented at a rate of 2 Hz, i.e. with a stimulus onset asynchrony of 500 ms, with no inter-stimulus interval. Images contained either a scene without any people (non-target) or a scene with a person holding a gun (target). A total number of 110 target images and 1346 non-target images were presented to each participant. Scenes in which a target appeared were also presented without the person in the non-target condition. All stimuli appeared within 6.5 degrees of center of the monitor. The goal of the task was to classify target images from non-target images. Participants responded to targets by either pressing a button or by silently counting. Single-trial detection was conducted from the session in which the subjects had only to count the number of target images. Electrophysiological recordings were digitally sampled at 1024 Hz from 64 scalp electrodes arranged in a 10-10 montage using a BioSemi Active Two system (Amsterdam, Netherlands). Impedances were kept below 25 $k\Omega$. External leads were placed on the outer canthus of both eyes and above and below the right orbital fossa to record EOG.

5. System architecture

Single-trial detection requires several steps that transform the signal information into different subspaces, from the raw EEG to the classifier output. Several approaches have been proposed in the literature, and the best methods include a deep hierarchical architecture. In shallow architectures, features are extracted from the EEG signal and are then given as input to a classifier. In deep architectures, the raw EEG signal is successively transformed into different subspaces through optimization procedures. The optimization steps can be independent, e.g., a first stage that maximizes the signal to noise ratio, then a second stage that minimizes the mean square error for the classification, or dependent, i.e., the first and second stages are linked together to minimize the error rate of single-trial detection. In deep architectures with independent steps, we can distinguish two categories based on the subspaces on which the signals are projected. With hierarchical linear discriminant analysis (H-LDA), each trial is decomposed into different windows, and a classifier is used for each window, and then the decisions from the different windows are combined to obtain the final decision. While the architecture of H-LDA has several stages, it does not belong to the regular deep architecture type as the first stage directly projects the information into the decision stage. The two stages correspond here to a multi-classifier system where different classifiers are used on different parts of the inputs, and another classifier combines the decisions. However, this process can be seen as a means to do both spatial and temporal filtering in a single step as the decision obtained from a window of the ERP signal merges the information in both space and time. If the width of the window is equivalent to a sampling point, then the resulting decision for each sampling point will correspond to spatial filtering (e.g. linear filtering with LDA). Another solution consists of setting all the parameters of the model (spatial filters and classifier) in a single step by using a convolutional neural networks [15]. With this type of approach using stochastic learning, the architecture of the network and the learning rates must be carefully chosen.

We consider $X \in \mathbb{R}^{N_t \times N_s}$ the recorded EEG signal, where N_t is the total number of sampling points, and N_s is the number of sensors. M is the number of different types of ERPs that are present in the signal. For two types of ERPs, e.g., target and non-target, $M = 2$. We denote by $\theta_j \in \mathbb{R}^{N_{C_j}}$ the vector containing the triggers of all the responses corresponding to the event of the class C_j, $1 \leq j \leq M$. Therefore, each trial i of the class C_j can be represented by a matrix $P_{i,j} \in \mathbb{R}^{N_{t1} \times N_s}$ where N_{t1} is the number of sampling points that include the effect of the stimulus on the brain activity after the stimulus onset at $\theta_j(i)$, $1 \leq i \leq N_{C_j}$. The recorded EEG signal can be defined as the addition of the evoked responses for each stimulus, i.e., the ERPs, and the noise.

In the raw EEG signal, the noise arises from various sources. The noise in the ERP, (i.e. signal that does not represent information about the experimental conditions), has four main origins: the ongoing brain activity not related to the task, the experimental design, the electrical activity produced by the subject outside of the brain (e.g. muscle artifacts), and activity produced by the environment (e.g. electrical line noise) . First, it is the EEG activity that does not correspond to a stimulus response. For instance, ongoing alpha waves can be a source of problem

if an ERP component has a strong activity in the alpha band. The second source of noise is the ERP variation across trials due to ongoing brain activity and differences across stimuli of the same class (e.g. different images corresponding to the same class of expected ERPs). This variation will be higher if there exists a large variation across stimuli in an experimental condition (e.g. brightness, contrast, possible meaning of the stimulus for the user). For instance, each possible item to select in the P300 speller is different, hence there should be theoretically as many target ERPs as items in the P300 speller. During RSVP tasks with natural images, each ERP will be different as each image can be different. The third type of noise is related to all the electrical activity that is not produced by the brain but present in the recorded EEG signal such as blinks and muscle activity. Typically blinks and other physiological artefacts can be removed from the EEG signal with independent component analysis (ICA) [42, 56]. The signal can be therefore decomposed as:

$$X = \sum_{j=1}^{M} \sum_{i=1}^{N_{C_j}} D_{i,j} P_{i,j} + H_0 \quad (1)$$

where the matrix $D_{i,j} \in \mathbb{R}^{N_t \times N_{t1}}$ is a Toeplitz matrix, $D_{i,j}$ has null values except for $D_{i,j}(\theta_j(i)-1+s, s) = 1, 1 \leq s \leq N_s$. Hence, the result of the product $D_{i,j}P_{i,j}$ represents a signal of size $N_t \times N_s$ that contains only the i^{th} ERP of the class C_j, and X represents the sum of all the evoked responses, with the noise H_0.

As there exists a common pattern between the evoked responses of a same class, each response can be decomposed into two components, the condition relevant ERP (Pc_j), and the difference across trials that can be attributed to the specificity of the current stimulus ($H_{i,j}$).

$$P_{i,j} = Pc_j + H_{i,j} \quad (2)$$

In the next sections, we denote by D_j the expression: $\sum_{i=1}^{N_{C_j}} D_{i,j}$, which represents the projection of the representative ERP of the class C_j on the different stimuli recorded in the signal X. The grand average response for the trials corresponding to the presentation of an event of the class C_j can be defined as follows:

$$\Sigma_j = \frac{1}{N_{C_j}} \sum_{i=1}^{N_{C_j}} P_{i,j} \quad (3)$$

Finally, the grand average difference between the events of the classes C_{j1} and C_{j2} can expressed by:

$$\Sigma_\Delta = \Sigma_{j1} - \Sigma_{j2} \quad (4)$$

6. Pre-processing

6.1. Temporal filtering

Temporal filtering can be achieved with linear filters such as Butterworth or Chebyshev. Decimation can be used to reduce the number of sampling points in the signal. Decimation has two effects: first to decrease the processing time of the different algorithms, second to decrease the number of features that have to be processed. To avoid aliasing, the decimation parameter must take into account the parameters of temporal filtering. For instance, a maximum cutoff frequency of $f_{sr}/2$ Hz can be set in a signal of a sampling rate f_{sr} Hz, yet it is often set to $f_{sr}/3$ Hz. In the following evaluations, the signal is downsampled to 64 Hz and bandpassed between 0.1 Hz and 21.33 Hz. In the literature, some methods use a lowpass filter at 10.66 Hz [50]. In [48], the signal is downsampled to 51.2 Hz, and bandpassed between 1 and 17 Hz. In [5], it was suggested that the best low cutoff frequency is 0.1 and the best high cutoff frequency is 15 Hz. The time segment that is used for the classification depends on the ERP components that appear between two conditions. In the case of target vs. non-target image classification, i.e. all the oddball paradigms with visual stimuli, the time segment includes all the signal poststimulus until around 800 ms. Hence, all the so-called P3 classifiers also include all the ERP components that appear before the P3, such as the N2.

6.2. Spatial filtering

Spatial filtering has a key role in single-trial detection as it enhances the signal and reduces the number of features. Spatial filtering techniques can be divided into several categories based on their purpose. Data driven approaches such as principal component analysis (PCA) and ICA [18] can be applied without any knowledge about the events/triggers corresponding to the stimuli/response onsets. Therefore, these methods do not exploit the particular differences between the brain responses evoked in one condition versus another condition. Spatial filters based on PCA determine filters that are uncorrelated and do not extract information related only to the discrimination of two types of signals. ICA provides spatial filters that can be particularly relevant for source localization in the brain. The sources of the signal are estimated by optimizing a criterion such as the negentropy [39, 40]. While ICA filters can help to remove the noise (e.g. eye blinks), they do not provide explicitly discriminant filters. In [38], no statistical difference in classification accuracies was found before and after using ICA. A strong assumption in spatial filtering methods in single-trial detection is that the differences observed in the EEG signal across trials within a class are not related to variation of the neural sources. Hence, a steady spatial distribution is expected over time for the generation of a particular type of ERP. With linear spatial filtering methods of a signal X, $U \in \mathbb{R}^{N_s \times N_f}$ corresponds to the set of N_f spatial filters, and XU is the obtained signal after spatial filtering. Linear discriminant spatial filtering methods are typically based on the maximization of a value such as the signal-to-signal plus noise ratio (SSNR). In such a case, the optimized spatial filters $\hat{U}$ are defined by:

$$\hat{U} = \underset{U}{\operatorname{argmax}}\, \mathrm{SSNR}(U) \quad (5)$$

with

$$\mathrm{SSNR}(U) = \frac{\mathrm{Tr}(U^T A U)}{\mathrm{Tr}(U^T B U)} \quad (6)$$

where Tr(.) denotes the trace operator, A represents an estimation of the signal power (e.g. the information relative to a specific ERP), and B represents the signal plus noise, or noise power,

respectively. If B is the identity matrix, i.e. there is no information about the noise or the other ERPs, and A is a covariance matrix, then the obtained spatial filters are equivalent to what can be achieved with principal component analysis. Within the xDAWN framework [72] (see Appendix A), the SSNR is evaluated with:

$$A = \widehat{(Xc_j)}^T (D_j)^T D_j \widehat{Xc_j}/N_t \quad (7)$$

$$B = X^T X/N_t \quad (8)$$

with $\widehat{(X_c^j)}$ being the least mean square, defined by:

$$\widehat{(Xc_j)} = ((D_j)^T D_j)^{-1} (D_j)^T X \quad (9)$$

7. Classification

7.1. Binary classification

Algorithms for classification typically require a training database containing a representative number of trials from each experimental condition. The number of required trials depends on what ERP components are expected to differ with respect to the experimental design. For regular ERP analysis, it is usually suggested, per condition, to use about 50 trials for large components such as the P3, and about 200 trials for the N2 wave, and up to 800 trials for small component such as the P1 [51]. The number of trials will depend on the experimental paradigm and its hypotheses. If the difference is assumed to occur at a late stage (e.g. the P3 versus no P3), then the number of required trials can be lower than if a difference of ERP across conditions should appear at P1. Overall, the number of trials will be strongly related to the difference of amplitude that is observed with the grand average waveform. If the difference of amplitude is small, then more trials will probably be required to model this difference.

The choice of the classifier is highly dependent to the number of available trials. In addition, for a person who wishes to exploit a classifier like a black box, some techniques are more interesting than others as they require less hyper-parameters. Despite the current strong interest in deep learning architecture for classification such as convolutional neural network, a large number of parameters must be determined in these models such as the number of neurons in each layer, the type of sigmoid functions, and the learning rate. After appropriate preprocessing, linear classifiers have been proven successful for binary classification. Linear discriminant analysis (LDA) and its variants, e.g., stepwise LDA, Bayesian LDA (BLDA) [35, 55] (see Appendix B), and support vector machines (with no particular kernel) have been successfully used. For instance, SVM has been effective in BCI competitions with appropriate features, showing state-of-the art performance [70].

A common problem with LDA based techniques is the estimation of an accurate covariance matrix in high dimensional space (sensor or spatial filters). In order to solve this problem, Blankertz et al. [4] have proposed regularization of LDA by shrinkage (covariance-regularized LDA). It involves the estimation of a parameter $\gamma \in [0, 1]$ such that $\gamma = 0$ represents the unregularized LDA, and $\gamma = 1$ assumes spherical covariance matrices. Another solution to deal with the problem of the covariance matrix in LDA is the creation of artificial training examples based on ERP characteristics such as the jitter effect. It increases the size of the database and has the advantage of being readily implemented for multiple classifiers. By adding deformed brain response signals in the training database, the number of trials is increased to estimate the covariance matrix, and the obtained classifier becomes invariant to small signal deformations that may occur in each trial. This technique requires apriori knowledge of the problem, i.e., the relationships between input features. An estimation of the variations across trials within a same class can directly provide the value of the parameters to use. In the case of artificial trials that are shifted in time (before and after the stimulus onset), it is strictly equivalent as providing additional shifted onsets in the list of events. In such a case, only the inputs (the list of triggers for each class) can be updated within the classification procedure. In the next section, we consider additional examples that are shifted in both directions with 31.25 ms, equivalent to two time points at 64 Hz.

7.2. Performance evaluation

A cross validation (CV) procedure is often used to assess the performance of a classifier. It is however worth noting that the type of cross validation can have a significant impact on the classification results and their interpretations [1]. CV is a model validation technique for assessing how the results of a classifier will generalize to a new independent data set. It is principally used for classification and prediction, and then the goal is to estimate how accurately a classifier will perform in practice. In a supervised classifier, a model is usually given a data set of labeled data on which training is run, and a data set of unlabeled data against which the model is tested. Exhaustive CV are typically not used because it is computationally expensive to learn and test a classifier on all possible ways to divide the original data set into a training and a validation set. Hence, non-exhaustive CV approaches are used because they do not compute all ways of splitting the original data set. Leave-one-out cross-validation involves using a single data point as the validation set, and the remaining data as the training set. This type of CV is often used when a database does not contain a sufficient number of trials, and it is better to dedicate as many trials as possible for training the classifier. On one hand, this may be inappropriate for the evaluation of BCI systems because the signals that are before and after the signal that is tested, are used during training, allowing the classifier to better capture the variability of the signal over time. Because the EEG signal is assumed to be non-stationary with fluctuations due to the subject's attention on the task, and the current subject's fatigue level, leave-one-out CV may not represent a realistic estimation of the classifier performance during the test. On the other hand, this approach allows the classifier to become tolerant to non-stationary characteristics of the EEG signal, and the classifier will be able to better capture the difference between two conditions.

Performance analysis is typically assessed by the area under the ROC curve (AUC) [28], due to the imbalanced class distribution. In addition, the AUC is a tool that only shows the potential performance of the method, and it does not represent what will happen during the test. Four databases should be ideally considered to properly evaluate a single-trial classification

system in a BCI setting: a training database to determine the parameters of the classifier, a validation database to determine the hyper-parameters of the classifier, a test database to determine the decision threshold, and finally an evaluation database to estimate the overall performance. The decision threshold can be selected to maximize the f-score in the training data-set, where the f-score is defined by:

$$\text{f-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (10)$$

$$\text{precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (11)$$

$$\text{recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (12)$$

where TP, FP, and FN corresponds to the number of true positive, false positive, and false negative decisions.

7.3. Information transfer rate

The performance of a BCI can be evaluated by the information transfer rate (ITR) [74] in bits per minute (bpm) defined by $ITR = \frac{60}{T} \cdot \vartheta$ where

$$\vartheta = log_2(N_{out}) + Plog_2(P) + (1-P)log_2(\tfrac{1-P}{N_{out}-1}) \quad (13)$$

and P being the probability of the good detection, i.e., the accuracy, N_{out} being the number of possible different outputs, and T being the time in seconds of recorded EEG signal that is required to take the decision among the N_{out} outputs.

This equation is not suited for the evaluation of single-trial detection because there is no balanced prior probability for the two classes. This is why the Nykopp definition of the ITR is more suitable:

$$\psi = \vartheta_0 - \vartheta_1 \quad (14)$$

$$\vartheta_0 = -\sum_{j=1}^{N_{out}} p(w_j) \cdot log_2(p(w_j)) \quad (15)$$

$$\vartheta_1 = -\sum_{i=1}^{N_{out}} \sum_{j=1}^{N_{out}} p(w_i) \cdot p(w_j|w_i) \cdot log_2(p(w_j|w_i)) \quad (16)$$

with $p(w_j|w_i)$ being the element (i, j) in the confusion matrix of the classification results.

7.4. Sensor selection

The goal of sensor selection is to select the most relevant sensors to extract information. Whereas spatial filters combine all the input channels into a set of new channels (i.e. from sensors to virtual sensors), sensor selection reduces the overall number of channels to select the most relevant ones. Backward elimination (BE) is typically used for sensor selection. In BE, the whole set of sensors is first used, and then the sensor that has the less negative impact on the classifier performance (AUC) is removed. In the evaluation, we consider the procedure where two sensors are removed at each stage.

8. Results

8.1. ERP analysis

Figure 3 depicts the grand-average waveforms of target and non-target trials. The ERPs for each sensor and all the trials are sorted based on the maximum value between 300 and 700 ms.

The topographic maps of the spatial distribution of the P3, obtained as the mean amplitude between 300 and 700 ms, is presented for each subject in the first and fourth row of Fig. 4. The spatial distribution obtained by the first component of the xDAWN approach is depicted in the second and fifth rows, while the corresponding spatial filter is represented in rows three and six. These figures show the relative similar distribution between the P3 spatial distribution and the distribution obtained through xDAWN. It is also worth noting the difference between spatial distribution and spatial filter as both have different meanings and use. Furthermore, as the spatial filters are set to maximize the SSNR, there is no control on the sign of the filters, i.e. if all the weights are multiplied by -1 it will not change their discriminant power. For this reason and to facilitate the comparisons across subjects, the weights of the spatial distributions and the spatial filters are set so the amplitude of Pz has the same sign as in the spatial distribution of the P3, and its sign drives the sign of the other channels.

The grand average waveforms corresponding to the presentation of target (bold line) and non-target stimuli with the four best spatial filters (SF_1 to SF_4) obtained with xDAWN with a time segment of 800 ms post stimulus are presented in Fig 5. The grand average waveform with the best spatial filter (SF_1) shows the significant contribution of the late ERP components for the detection of target vs. non-target trials.

8.2. Single-trial performance

The performance over time, from -125 ms to 1000 ms, using an LDA classifier is depicted in Fig. 6. In this figure an LDA classifier is used for different time segments of size 31.25 ms, every 31.25 ms. The best performance is achieved at 500 ms, with an average AUC of 0.777 ± 0.084, which is significantly above chance level. The AUC stays around 0.5 until 200 ms where the AUC increases to reach a peak at 500 ms, and then decreases progressively. This analysis confirms that the most discriminant information for single-trial detection in this data is in a late ERP component (P3), and that early components have a significant impact on the classification. Particularly, the p-value of pairwise comparisons between the AUC obtained from each subject and chance level (AUC=0.5) is 0.04 at 62.5 ms, and 0.02 at 93.75 ms. These results clearly indicate that while single-trial detection cannot be achieved with early ERP component because the performance is very low, they have still a significant discriminant contribution. The red circles in the figure represents the time point where a significant difference is obtained with a Wilcoxon signed-rank test, after a Bonferoni correction. It shows the possibility of discriminating between target and non-target above chance level as early as 125 ms, with an AUC=0.54. It is worth mentioning that at 218 ms, the average AUC is 0.58, yet it is not significantly above chance when compared across subjects.

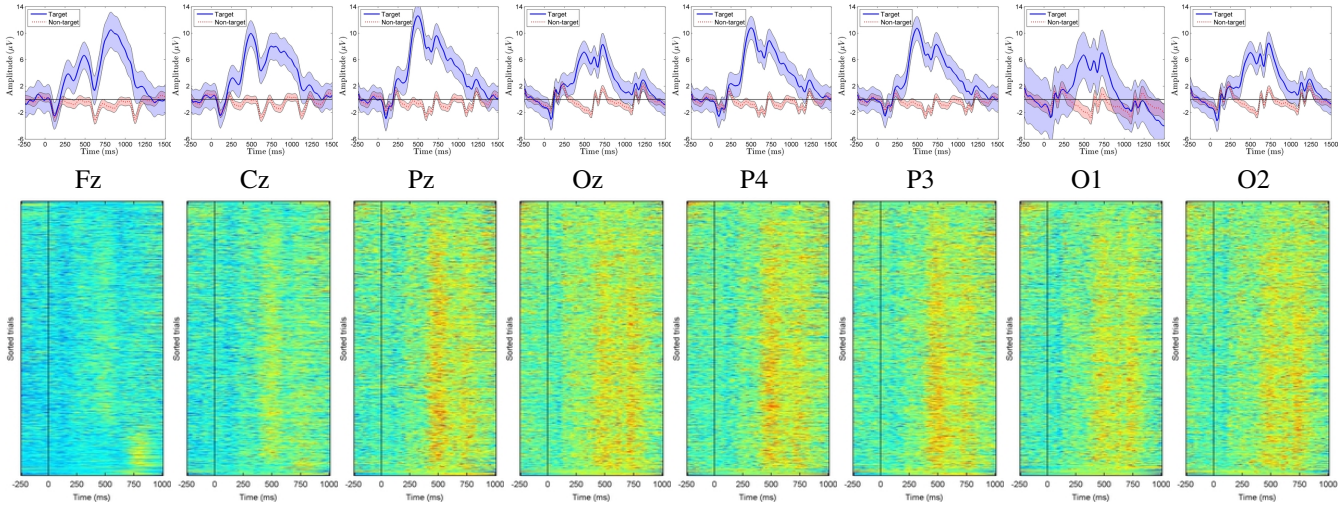


Fig. 3. Grand average waveform corresponding to the presentation of target (bold line) and non-target stimuli with the electrodes Fz, Cz, Pz, Oz, P4, P3, O1 and O2, with the corresponding sorted trials. The areas correspond to envelope +/- the standard error across subjects of the ERP waveform.

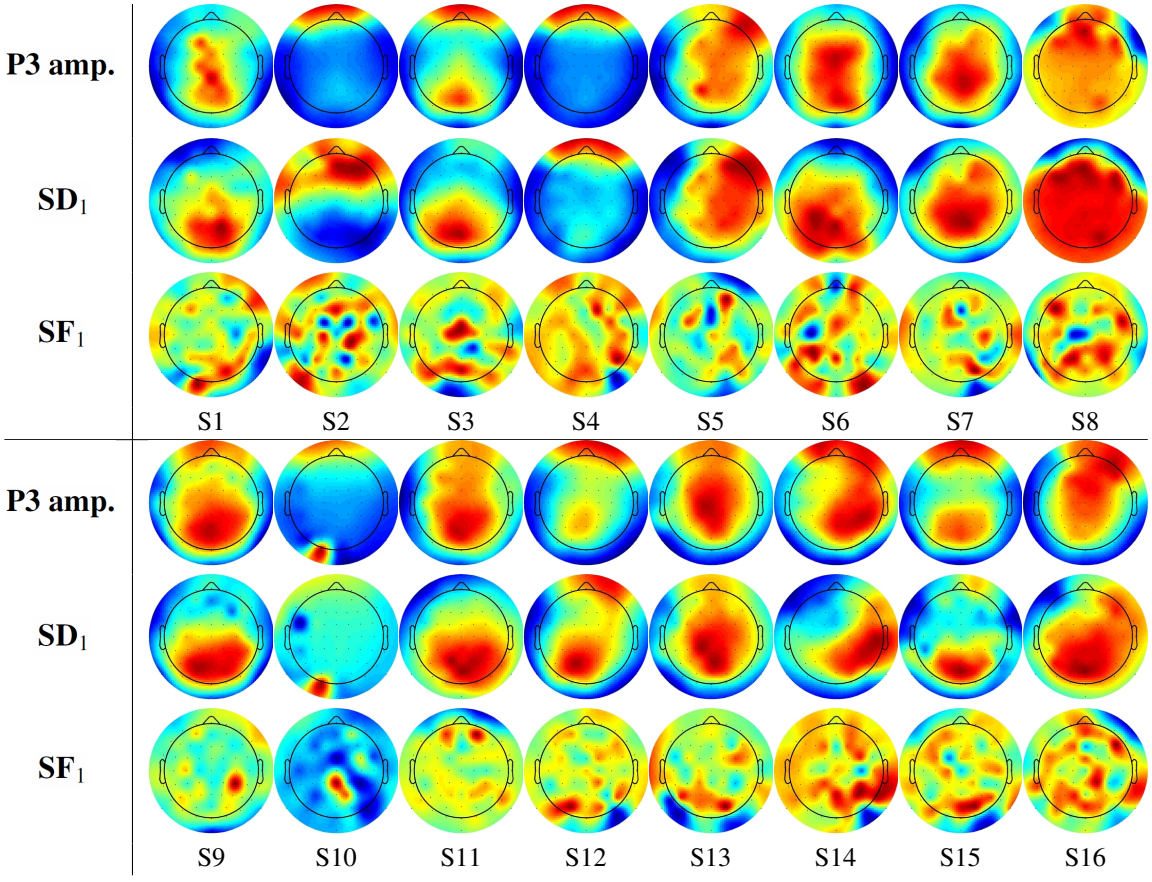


Fig. 4. Topoplots representing the spatial distribution of the P3 [300-700 ms], the discriminant spatial distribution (SD) obtained through xDAWN (first best filter), and its corresponding spatial filter (SF).

The AUC for different classifiers is depicted in Fig 7, 8, and 9. The value of the mean AUC across subjects is given in Table 1 and 2. These results correspond to an evaluation with a 5-fold cross-validation procedure, using a time segment of 800 ms starting after the stimulus onset. H-LDA corresponds to Hierarchical LDA with the use of 4 segments (every 200 ms) which are processed with LDA, then the LDA classifier of each time segment is used as a spatial filter, the final classifier being LDA. The same principle is used for H-BLDA and H-swLDA, where each segment is classified with BLDA (sw-LDA for H-swLDA), and the final classifier is also BLDA (sw-LDA for H-swLDA). With the shift option, trials shifted by 2 time points before and after the stimulus onsets are added to the training database. The best performance is obtained with xDAWN+BLDA+Shift with

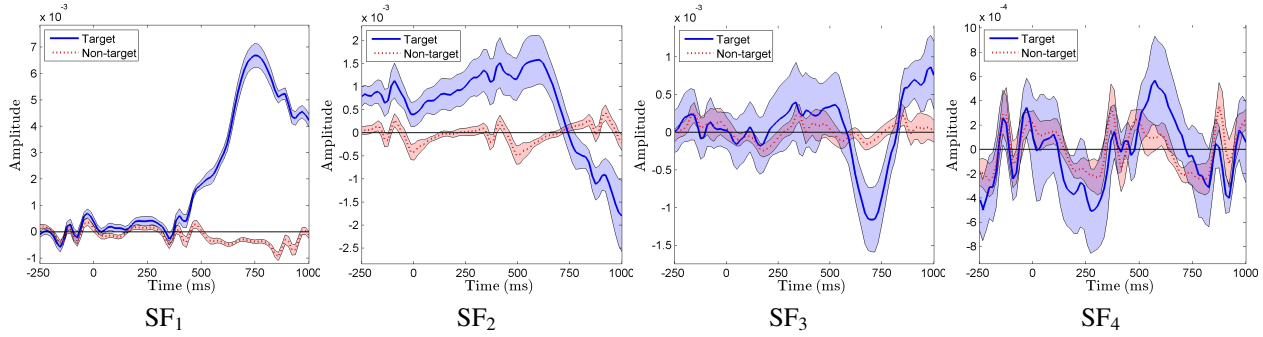


Fig. 5. Grand average waveform corresponding to the presentation of target (bold line) and non-target stimuli with the four best spatial filters obtained with xDAWN (0-800ms). The areas correspond to envelope +/- the standard error across subjects of the filtered ERP waveform.

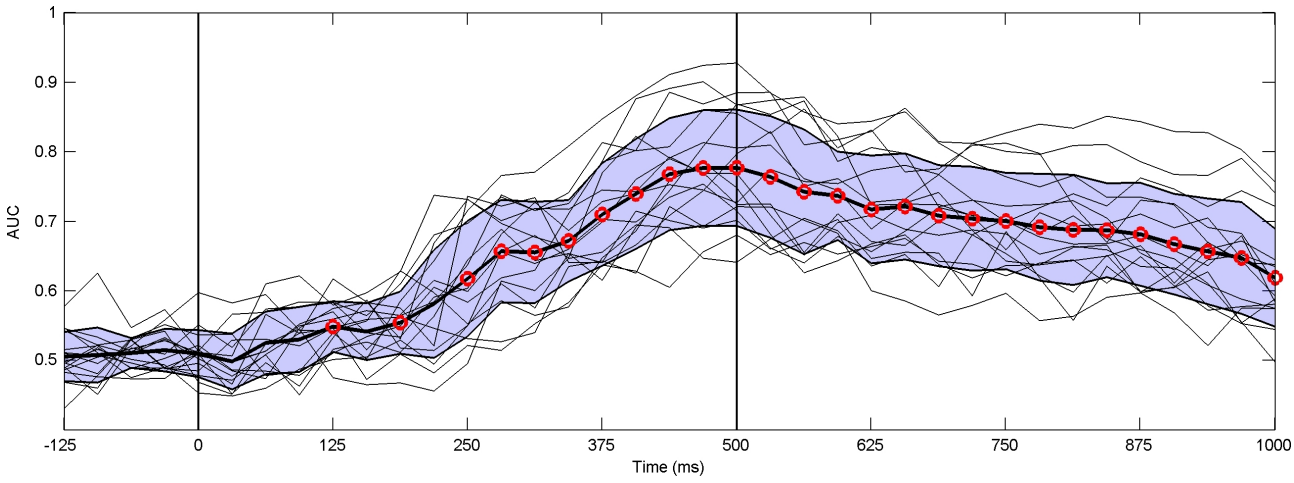


Fig. 6. AUC performance over time with an LDA classifier using a window of 31.25 ms. The bold line represents the average AUC across subjects and the blue envelope represents ± the standard deviation.

an AUC=0.954 ± 0.030 across subjects. The results are almost similar with xDAWN+BLDA only (AUC=0.953±0.032) (no statistical difference). The results confirm the efficiency of spatial filtering (xDAWN and the LDA/BLDA/swLDA based spatial filters) for both LDA, BLDA, and swLDA. Pairwise comparisons confirm the superiority of xDAWN+BLDA over the other methods. Classifiers without spatial filtering for reducing the number of features (LDA and SVM) could not provide a decision above chance level. The estimated ITR for xDAWN+BLDA+Shift is 40.62 ± 0.63 bits per minute, after using a decision threshold maximizing the f-score.

The performance in relation to the sampling rate and the lowpass filtering option is depicted in Fig. 10. In this evaluation, the AUC corresponds to a cross-validation procedure on the whole recorded signal by using xDAWN spatial filters with a BLDA classifier and artificial shifted trials to enrich the training database. The results indicate that it is possible to achieve an above chance performance with all the different signals and with less than 10 trials representing a target. With both the minimum and maximum number of trials, the best performance is obtained with a signal at 64 Hz and a lowpass filter set to 21.33 Hz, with an AUC of 0.770 with 5 target trials, and 0.934 with 55 target trials. Decreasing the sampling rate to 32 Hz and therefore decreasing the number of features does not improve single-trial performance. As the number of available trials for training can have a significant impact on the performance, we compare the performance between xDAWN+BLDA and xDAWN+BLDA+Shift. Fig. 11 presents the AUC as a function of the number of targets in the training database. First, the results indicate that at least 50 trials corresponding to a target must be present in order to reach a reliable AUC. Second, the addition of shifted trials allows for a significant increase in performance when the number of trials is low. These results explain that the requirement of a high number of trials is the temporal variability of the brain evoked responses.

8.3. Sensor selection evaluation

The performance obtained with different subsets of sensor is depicted in Fig. 13. The AUC increases in relation to the number of sensors. Yet, with only 2 sensors that are specific to each subject, it is possible to obtain an average AUC of 0.867. With the whole set of 64 sensors, the average AUC is 0.923. These results indicate that a low number of sensors, which are specific to an individual, can be enough for single-trial detection. The rank of the sensors is displayed for each subject in Fig. 12.

9. Discussion

Single-trial detection of ERPs in EEG/MEG signals is a key element in brain-computer interface systems. Single-trial detection requires preprocessing using adapted techniques in relation

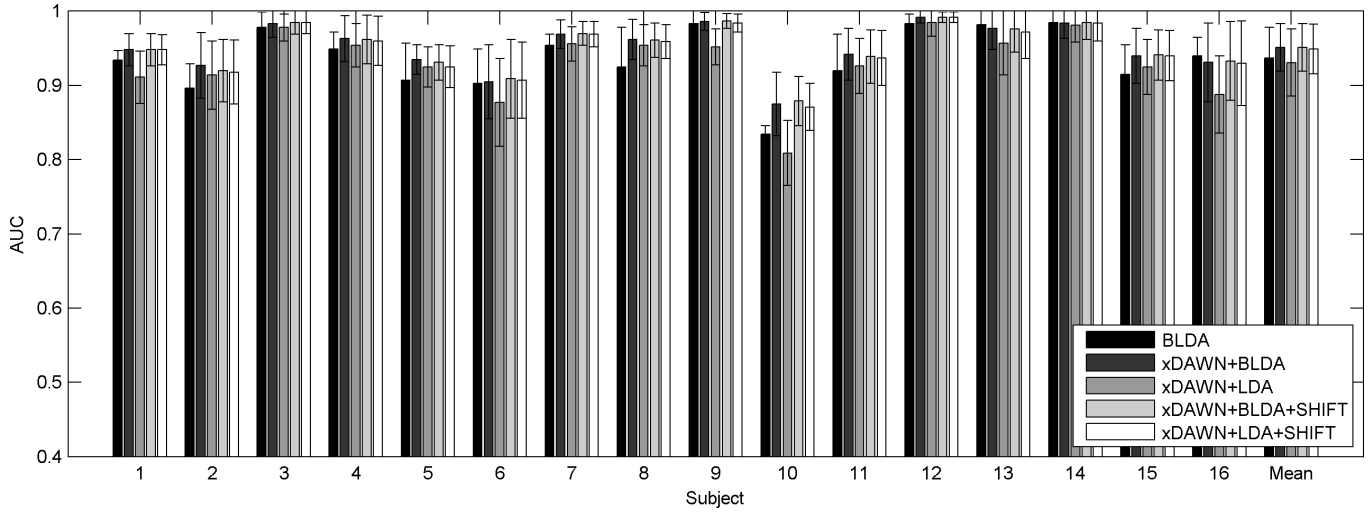


Fig. 7. AUC performance for different methods (BLDA, xDAWN+BLDA, xDAWN+LDA, xDAWN+BLDA+Shift, and xDAWN+LDA+Shift).

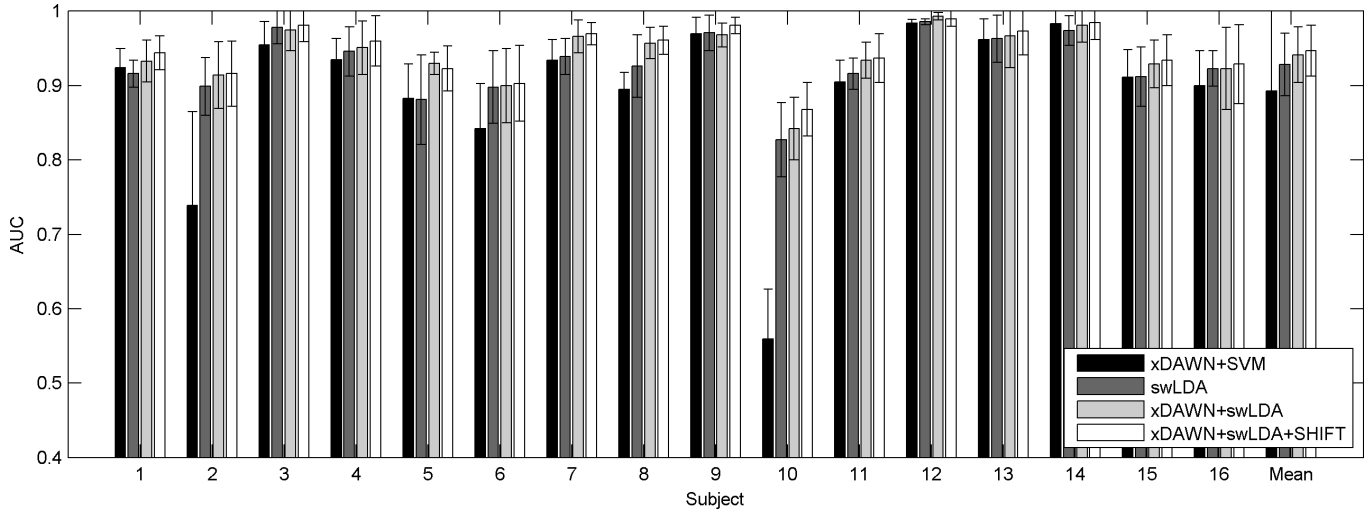


Fig. 8. AUC performance for different methods (xDAWN+SVM, swLDA, xDAWN+swLDA, and xDAWN+swLDA+Shift).

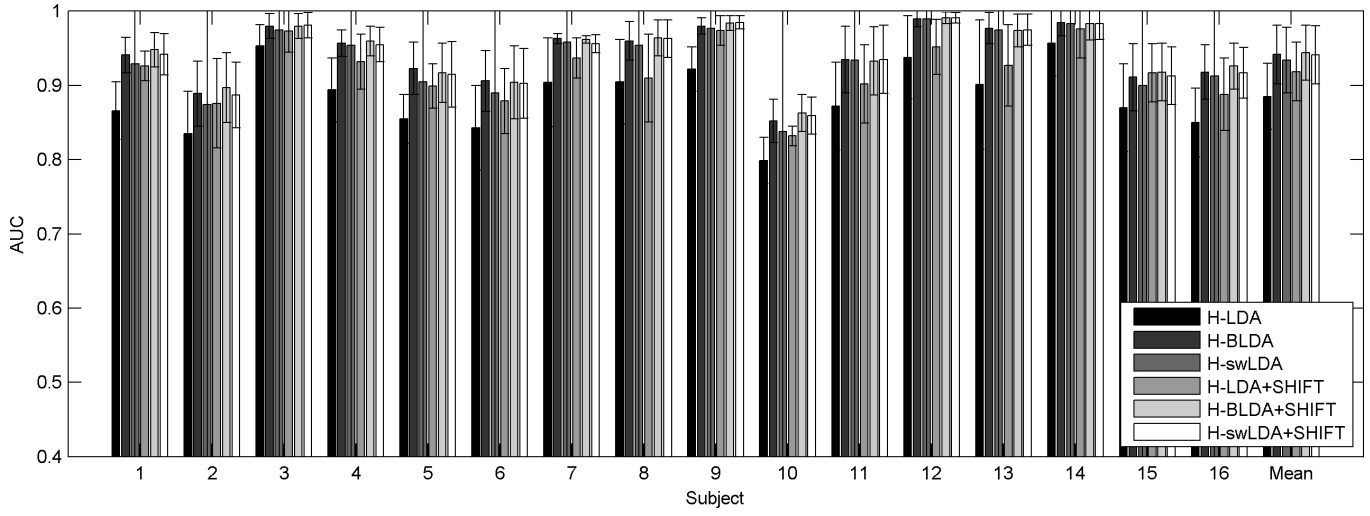


Fig. 9. AUC performance for different methods (H-LDA, H-BLDA, H-swLDA, H-LDA+Shift, H-BLDA+Shift, H-swLDA+Shift).

Table 1. Single-trial performance for linear classifiers, with and without spatial filtering.

	BLDA	BLDA	LDA	BLDA	LDA	SVM	swLDA	swLDA	swLDA
xDAWN		x	x	x	x	x		x	x
Shift				x	x				x
mean AUC	0.937	0.953	0.930	0.954	0.949	0.893	0.928	0.941	0.947
sd AUC	0.042	0.032	0.045	0.032	0.033	0.108	0.042	0.037	0.034

Table 2. Single-trial performance for hierarchical approaches.

	H-LDA	H-BLDA	H-swLDA	H-LDA	H-BLDA	H-swLDA
Shift				x	x	x
mean AUC	0.885	0.942	0.934	0.919	0.944	0.941
sd AUC	0.044	0.039	0.044	0.040	0.037	0.039

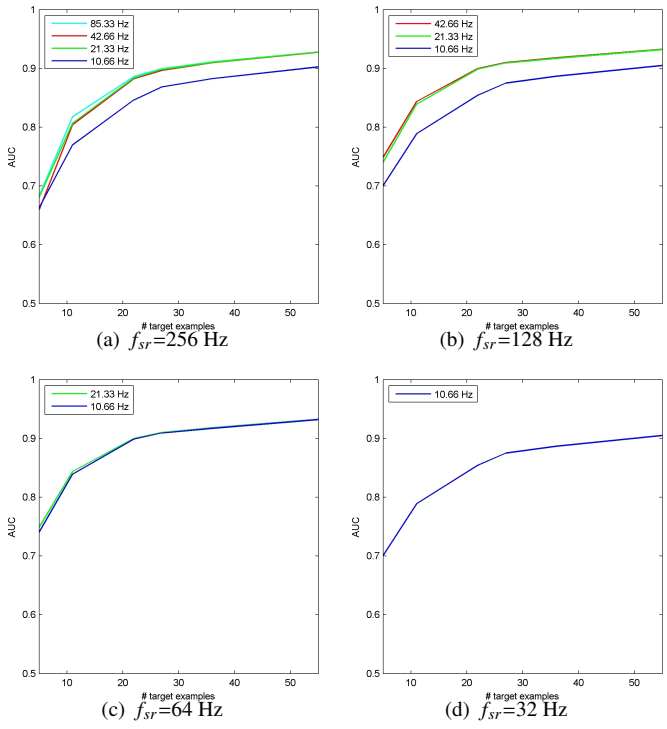


Fig. 10. AUC (xDAWN+BLDA+artificial shifted trials) for signals of different sampling rates (256, 128, 64 and 32 Hz) and lowpass filtering options in relation to the number of targets for training.

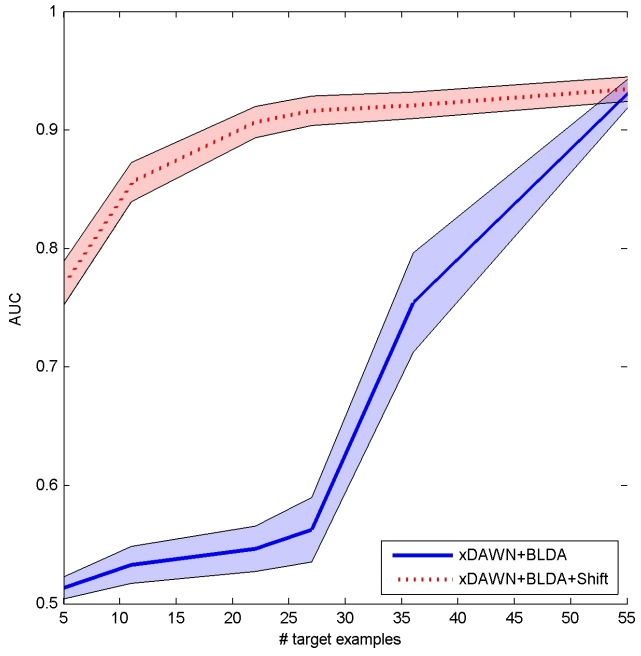


Fig. 11. AUC performance as a function of the number of trials corresponding to a target. The envelopes represent the standard error across subjects.

to the expected brain responses. The different preprocessing steps include bandpass filtering and downsampling. The present results indicate that bandpass filtering between 0.1 and 21.33 Hz, with downsampling the signal to 64 Hz provide an efficient performance. Downsampling the signal is a fast and simple way to reduce the number of time points before classification. In addition, we have shown that spatial filtering is a key component for both enhancing the signal, and decreasing the number of features. This preprocessing step is particularly relevant if there is no prior information about the spatial distribution of a brain response, i.e., if all the electrodes are selected to cover the whole scalp. Finally, if spatial filtering is used before classification, the choice of the classifier is not critical. In addition, increasing the number of trials through the addition of shifted trials (or shifted stimuli onsets) can provide an additional increase of performance when there exists a large difference of latency across trials. Methods that are well suited for motor-imagery detection such as common spatial patterns are not optimal for the detection of event-related potentials. It can be sufficient to only use a classifier with bandpass filtered EEG as input when the number of trials is large enough (superior to 50), and the type of ERP component to detect is clearly identified; however, the analysis of the spatial distribution through spatial filtering can reduce the number of features and enhance the signal given as input to a classifier.

Single-trial detection cannot be applied successfully on any ERP. The same basic principles that apply for ERP analysis should be applied for single-trial detection. It is critical to have prior information about the type of ERP components that will be present. Key characteristics such as the inter stimulus interval, the stimulus onset asynchrony, the target-to-target interval, and the target probability will have an effect on the amplitude of ERP components such as the P3, and they will have an effect on single-trial detection [10, 15], and latency [58]. In the present

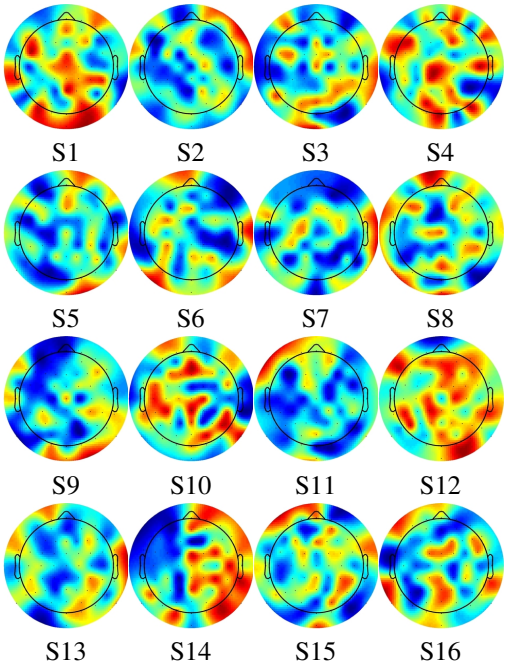


Fig. 12. Sensor ranking for each subject (red represents a high rank, blue represents a low rank).

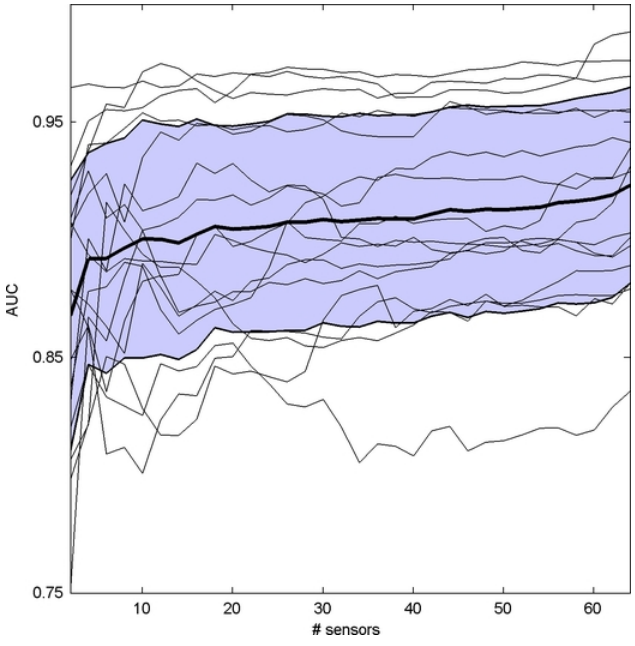


Fig. 13. AUC performance across different number of sensors.

study, we have shown that it is possible to obtain above average classification performance with a limited number of trials corresponding to the presentation of a target. This performance indicates that single-trial detection can be used as a reliable tool to determine the presence of a large ERP component when the lowest number of trials in one class is low (e.g. about 20 trials).

The choice of the architecture and the classifier remains difficult, and depends mainly on the number of available trials. Whereas linear methods stay popular as they are efficient and easy to implement [63], it has been shown that Gaussian SVM and the neural networks can outperform linear classifiers [48]. With a large number of trials, non-linear classifiers and deep architecture should be able to capture high level features related to the variability of the signals over time. However, because acquiring EEG signal for a specific subject is time consuming, and the EEG signal has non-stationary properties, tracking changes or adapting the system over time may be more judicious in BCI settings. The choice of the method should be driven by the amount of available data for training, and to what extent the signal can change over time. Linear classifiers (e.g. linear SVM, LDA) are often favored over non-linear ones due to their simplicity, which helps to prevent overfitting on the noisy and limited signals [61, 62]. With that said non-linear classifiers have been successfully utilized in BCI [11, 20, 61]. Complex neuroimaging signals may be intrinsically non-linear and require non-linear classifiers, therefore, it is important to consider the whole preprocessing pipeline that is used before the classification. Domain knowledge can help reduce and select features with non-linear or linear techniques, and then apply a linear classifier. Because EEG signal is typically subject specific, the amount of trials can be scarce, and therefore efficient model regularization is difficult due to the extensive hyperparameter space. However, it is suggested that the Gaussian processes classifier is capable of learning relevant non-linearities in data while tuning the kernel hyperparameters without requiring extensive cross-validation and without overfitting [43]. The same principles apply to other neuroimaging data such as fMRI where linear classifiers have been mainly used (e.g. to decode object category from voxel activity in ventral temporal cortex [32], to decode the orientation of a grating rendered subjectively invisible by a mask [33]).

The methods presented in this paper, while proven efficient on single-trial detection of ERPs in EEG, can also be used with magnetoencephalography (MEG) [8, 13]. Single-trial detection of event-related fields can be achieved with the same methods. While the results presented in this article are based on responses to visual stimuli, the same methods can be used for ERPs based on auditory stimuli. Moreover, spatial filtering is particularly significant when the number of sensors is high (256 in high density EEG, 306 with MEG). Using a specific time segment that isolates an ERP component, single-trial detection performance can be useful for determining which components are influenced by an experimental condition (e.g. target vs. non-target in an oddball paradigm). Classification over different time segments can show temporal differences in processing stages between conditions.

Current methods for single-trial detection are still limited in their ability to capture and adjust their output based on the global state of the individual. First, they do not take into account fluctuations of brain responses over time that can have a key impact on the accuracy. Fatigue, mind wandering, and change of task difficulty can result in different brain states that impact the spatial distribution of the evoked responses over time. New models should be proposed to account for the non-stationary distributions of the features contained in this variability. Second, the difference across trials is not properly taken into account as current methods do not include mechanisms to shift ERPs based on the grand average model, without including informa-

tion about the reaction time. Moreover, popular techniques only take advantage of the presence of a large ERP component for its detection, and the binary classification stays at the level of "present" versus "absent". Single-trial classification methods could benefit by focusing beyond the detection of a late ERP component. Brain decoding techniques can be improved by increasing the transfer rate of information complexity; i.e. not only the presence of a particular stimulus, but its class, and its spatial location. For example, a multi-class classification problem may focus on distinguishing between rare targets, rare non-targets, and frequent non-targets where the two rare stimuli produce similar overlapping components (i.e. P3a, P3b) thereby increasing the number of stimuli type to detect [14]. Second, early components, e.g. P1/N1, are not currently used because of their low signal-to-noise ratio. Single-trial detection of the P1/N1 can provide information about the spatial location of the stimulus in relation to where the subject is paying attention [53].

10. Conclusion

In this paper, a comprehensive description of an automatic pipeline for single-trial detection of event-related potentials has been presented. The proposed strategy has been proven successful on a binary classification problem during a rapid serial visual presentation task. More particularly, spatial filters based on the xDAWN framework followed by a linear classifier trained with data that include artificial shifted trials provide the best performance. Furthermore, a key interest in the approach is the reproducibility of the results as the presented methods have a limited number of hyperparameters, and do not use any stochastic or randomized approach.

Acknowledgment

This research was supported by the Institute for Collaborative Biotechnologies through contract W911NF-09-D-0001 from the U.S. Army Research Office. H.C would like to thank the NI Functional Brain Mapping Facility project (1303/101154803) funded by InvestNI and Ulster University.

Appendix A. xDAWN

The xDAWN spatial filtering technique [71, 72] is described as follows. The enhanced signal $XU \in \mathbb{R}^{N_t \times N_s}$ is composed of three terms: the ERP responses on a target class (D_1A_1), a response common to all stimuli, and non-targets confound (D_2A_2), and the residual noise (H), that are all filtered spatially with $U \in \mathbb{R}^{N_s \times N_f}$, N_f being the number of spatial filters, $1 \leq N_f \leq N_s$.

$$XU = (D_1A_1 + D_2A_2 + H)U. \tag{A.1}$$

where $\{D_1, D_2\} \in \mathbb{R}^{N_t \times N_1}$ are two Toeplitz matrices, N_1 is the number of sampling points representing the target and superimposed evoked potentials, and $H \in \mathbb{R}^{N_t \times N_s}$. The spatial filters $\hat{U}$ maximize the SSNR defined by:

$$\mathrm{SSNR}(U) = \frac{\mathrm{Tr}(U^T \hat{A}_1^T D_1^T D_1 \hat{A}_1 U)}{Tr(U^T X^T X U)} \tag{A.2}$$

where $\hat{A_1} \in \mathbb{R}^{N_1 \times N_s}$ represents the least mean square estimation of A_1:

$$\hat{A} = \begin{bmatrix} \hat{A_1} \\ \hat{A_2} \end{bmatrix} = ([D_1; D_2]^T [D_1; D_2])^{-1} [D_1; D_2]^T X \tag{A.3}$$

where $[D_1; D_2] \in \mathbb{R}^{N_t \times 2N_1}$ is obtained by concatenation of D_1 and D_2. In the SSNR definition, $\hat{A}_1$ is replaced by $B_1^T X$ where B_1^T is a part of the least mean square estimation; a QR decomposition is applied on $D_1 = Q_1 R_1$ and $X = Q_x R_x$. Then, a singular value decomposition (SVD) of $R_1 B_1^T Q_x$ provides $\Phi \Lambda \Psi^T$, where Φ and Ψ are two unitary matrices, and Λ is a diagonal matrix with non-negative diagonal elements in decreasing order. Finally, the spatial filters are $\hat{U} = R_x^{-1} \Psi$, ordered by decreasing order of relevance impact, and the first N_f filters are used.

Appendix B. BLDA

The BLDA technique [55] provides a set of weights $w \in \mathbb{R}^{m+1}$, m being in the number of features, that are defined as follows. We denote by N^+ and N^-, the number of examples belonging to class C_1 and class C_2, respectively, with $n = N^+ + N^-$, the total number of examples. We set the ground truth y as $y(i) = n/N^+$ if the example $x(i)$ belongs to the class C_1, $y(i) = n/N^-$ if $x(i)$ belongs to the class C_2, $1 \leq i \leq n$. The matrix $X \in \mathbb{R}^{m+1} \times \mathbb{R}^n$ contains the information relative to the n examples:

$$X = \begin{pmatrix} x_1(1) & \dots & x_1(n) \\ \vdots & \ddots & \vdots \\ x_m(1) & \dots & x_m(n) \\ 1 & \dots & 1 \end{pmatrix} \tag{B.1}$$

The classifier output $\hat{y}(i)$ for the example $x(i)$ is defined by:

$$\hat{y}(i) = \sum_{j=1}^{m+1} w(j) \cdot x(j, i) \tag{B.2}$$

The overall error is estimated by:

$$\mathrm{err} = \sum_{i=1}^{n} (y(i) - \hat{y}(i))^2 \tag{B.3}$$

The parameters of the classifier, w, are obtained by the following equation:

$$w = \beta \cdot V \cdot u_2 \tag{B.4}$$

where $V \in \mathbb{R}^{(m+1)^2}$ and $\lambda \in \mathbb{R}^{m+1}$ correspond to the Eigen vectors and the Eigen values of XX^T, respectively, and $u_1 = V^T X y'$. Then, we define u_2 as

$$u_2(i) = \begin{cases} \frac{u_1(j)}{\beta \lambda_j + \alpha} & \text{if } (j \leq m) \\ \frac{u_1(j)}{\beta \lambda_j + b_\alpha} & \text{if } (j = m + 1) \end{cases} \tag{B.5}$$

Then, γ, α, and β are defined by:

$$\gamma = \frac{\beta \lambda_{m+1}}{\beta \lambda_{m+1} + b_\alpha} + \sum_{j=1}^{m} \frac{\beta \lambda_j}{\beta \lambda_j + \alpha} \tag{B.6}$$

$$\alpha = \frac{\gamma}{\sum_{j=1}^{m+1} w(j)^2} \tag{B.7}$$

$$\beta = \frac{n - \gamma}{\mathrm{err}} \tag{B.8}$$

The estimation of γ, α, and β is obtained with the Algorithm 1.

Algorithm 1 BLDA.

$(\alpha_1, \beta_1) \leftarrow (25, 1), (\Delta\alpha, \Delta\beta) \leftarrow (\infty, \infty)$
$t \leftarrow 1, t_{max} \leftarrow 100, \epsilon \leftarrow 10e-4, b_\alpha \leftarrow 10e-8$
while $((\Delta\alpha > \epsilon) \,||\, (\Delta\beta > \epsilon))$ & $(t < t_{max})$ **do**
  $(\alpha_{t-1}, \beta_{t-1}) \leftarrow (\alpha_t, \beta_t)$
  Compute w, γ, (α_t, β_t)
  $(\Delta\alpha, \Delta\beta) \leftarrow (|\alpha_t - \alpha_{t-1}|, |\beta_t - \beta_{t-1}|)$
  $t \leftarrow t + 1$
return w

References

[1] Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Statist. Surv. 4, 40–79.

[2] Awni, H., Norton, J.J., Umunna, S., Federmeier, K.D., Bretl, T., 2013. Towards a brain computer interface based on the N2pc event-related potential, in: Proc. of 6th Int. Neural Eng. IEEE/EMBS Conf., pp. 1021–1024.

[3] Bigdely-Shamlo, N., Vankov, A., Ramirez, R.R., Makeig, S., 2008. Brain activity-based image classification from rapid serial visual presentation. IEEE Trans. on Neural Systems and Rehab. Eng. 16, 432–41.

[4] Blankertz, B., Lemm, S., Treder, M.S., Haufe, S., R., M.K., 2011. Single-trial analysis and classification of ERP components - a tutorial. Neuroimage 56, 814–825.

[5] Bougrain, L., Saavedra, C., Ranta, R., 2012. Finally, what is the best filter for P300 detection?, in: TOBI Workshop III- Tools for Brain-Computer Interaction - 2012, pp. 1–2.

[6] Cai, B., Xiao, S., Jiang, L., Wang, Y., Zheng, X., 2013. A rapid face recognition BCI system using single-trial ERP, in: Proc. of the 6th Int. Neural Eng. IEEE/EMBS Conf., pp. 89–92.

[7] Cecotti, H., 2011. Spelling with non-invasive brain-computer interfaces - current and future trends. J. Physiology-Paris 105, 106–114.

[8] Cecotti, H., 2015a. Single-trial detection with magnetoencephalography during a dual rapid serial visual presentation task. IEEE trans. Biomed. Eng. , 1–8.

[9] Cecotti, H., 2015b. Toward shift invariant detection of event-related potentials in non-invasive brain-computer interface. Pattern Recognition Letters 66, 127–134.

[10] Cecotti, H., Eckstein, M.P., Giesbrecht, B., 2014. Single-trial classification of neural responses evoked in rapid serial visual presentation: Effects of stimulus onset asynchrony and stimulus repetition. 36nd Int. IEEE Conf. of the Eng. in Medicine and Biology Soc. .

[11] Cecotti, H., Gräser, A., 2011. Convolutional neural networks for P300 detection with application to brain-computer interfaces. IEEE Trans. Pattern Analysis and Machine Intelligence 33, 433–445.

[12] Cecotti, H., Marathe, A., Ries, A.J., 2015. Optimization of single-trial detection of event-related potentials through artificial trials. IEEE Trans. Biomed. Eng. , 1–7.

[13] Cecotti, H., Prasad, G., 2015. Single-trial detection of realistic images with magnetoencephalography, in: Proc. of the Int. Joint Conf. on Neural Networks, pp. 1–6.

[14] Cecotti, H., Ries, A., Eckstein, M., Giesbrecht, B., 2012. Multiclass classification of single trial evoked eeg responses, in: Proc. of the 34nd Int. IEEE EMBC, pp. 1719–22.

[15] Cecotti, H., Sato-Reinhold, J., Sy, J.L., Elliott, J.C., Eckstein, M.P., Giesbrecht, B., 2011. Impact of target probability on single-trial EEG target detection in a difficult rapid serial visual presentation paradigm task. 33nd Int. IEEE Conf. of the Eng. in Medicine and Biology Soc. .

[16] Childers, D.G., Perry, N.W., Fischler, I.A., Boaz, T., Arroyo, A.A., 1987. Event-related potentials: a critical review of methods for single-trial detection. Crit Rev Biomed. Eng. 14, 185–200.

[17] Chun, M.M., Potter, C.M., 1995. A two-stage model for multiple target detection in rapid serial visual presentation. J. Exp. Psy. Human Perception and Performance 21, 109–127.

[18] Comon, P., 1994. Independent component analysis: a new concept? Signal Processing 36, 287–314.

[19] Croft, R.J., Gonsalvez, C.J., Gabriel, C., Barry, R.J., 2003. Target-to-target interval versus probability effects on P300 in one-and two-tone tasks. Psychophysiology 40, 322–328.

[20] Davatzikos, C., Ruparel, K., Fan, Y., Shen, D., Acharyya, M., Loughead, J., Gur, R., Langleben, D.D., 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. NeuroImage 28, 663–668.

[21] Delorme, A., Miyakoshi, M., Jung, T.P., Makeig, S., 2015. Grand average ERP-image plotting and statistics: A method for comparing variability in event-related single-trial EEG activities across subjects and conditions. J. Neuroscience Methods 250, 3–6.

[22] Dien, J., Spencer, K.M., Donchin, E., 2004. Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. Psychophysiology 41, 665–78.

[23] Donchin, E., 1966. A multivariate approach to the analysis of average evoked potentials. IEEE Trans. Biomed. Eng. 13, 131–139.

[24] Donchin, E., 1969. Discriminant analysis in average evoked response studies: The study of single trial data. Electroencephalography and Clinical Neurophysiology 27, 311–314.

[25] Donchin, E., Spencer, K.M., Wijesinghe, R., 2000. The mental prosthesis: assessing the speed of a P300-based brain-computer interface. IEEE Trans Rehabil Eng. 8, 174–9.

[26] Donnerer, M., Steed, A., 2010. Using a P300 brain-computer interface in an immersive virtual environment. Presence: Teleoperators and Virtual Environments 19, 12–24.

[27] Farwell, L., Donchin, E., 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr. Clin. Neurophysiol. 70, 510–523.

[28] Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters 27, 861–874.

[29] Folstein, J.R., Van Petten, C., 2011. After the P3: Late executive processes in stimulus categorization. Psychophysiology 48, 825–841.

[30] Gerson, A., Parra, L., Sajda, P., 2006. Cortically-coupled computer vision for rapid image search. IEEE Trans. Neural Syst. Rehabil. Eng. 14, 174–179.

[31] Guger, C., Daban, S., Sellers, E., Holznera, C., Krausza, G., Carabalonac, R., Gramaticac, F., Edlinger, G., 2009. How many people are able to control a P300-based brain.computer interface (BCI)? Neuroscience Letters 462, 94–98.

[32] Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.

[33] Haynes, J., Rees, G., 2005. Predicting the stream of consciousness from activity in human visual cortex. Curr. Biol. 15, 1301–1307.

[34] Hild, K.E., Kurimo, M., Calhoun, V.D., 2010. The sixth annual MLSP competition, 2010. IEEE Int. Workshop on Machine Learning for Signal Processing , 107–111.

[35] Hoffmann, U., Vesin, J., Diserens, K., Ebrahimi, T., 2008. An efficient P300-based brain-computer interface for disabled subjects. J. of Neuroscience Methods 167, 115–125.

[36] Hong, B., Guo, F., Liu, T., Gao, X., Gao, S., 2009. N200-speller using motion-onset visual response. Clinicial Neurophysiology 120, 1658–1666.

[37] Huang, Y., Erdogmus, D., Pavel, M., Mathan, S., Hild, K.E., 2011. A framework for visual image search using single-trial brain responses. Neurocomputing 74, 2041–2051.

[38] Hwang, H.J., Ferreria, V.Y., Ulrich, D., Kilic, T., Chatziliadis, X., Blankertz, B., Treder, M., 2015. A gaze independent brain-computer interface based on visual stimulation through closed eyelids. Scientific Reports 5, 1–11.

[39] Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. on Neural Networks 10, 626–634.

[40] Hyvärinen, A., Oja, E., 2000. Independent component analysis: Algorithms and applications. Neural Networks 13, 411–430.

[41] Johnson, R., 1988. The amplitude of the P300 component of the event-related potentials: review and synthesis, in: editors, A.P.J.J.C.M. (Ed.), Advances in psychophysiology, pp. 69–137.

[42] Jung, T.P., Humphries, C., Lee, T.W., Makeig, S., J., M.M., Iragui, V., Sejnowski, T.J., 1998. Extended ICA removes artifacts from electroencephalographic recordings, in: Advances in Neural Information Processing Systems, pp. 894–900.

[43] Kauppi, J.P., Kandemir, M., Saarinen, V.M., Hirvenkari, L., Parkkonen, L., Klami, A., Hari, R., Kaski, S., 2015. Towards brain-activity-controlled information retrieval: Decoding image relevance from MEG signals. NeuroImage 112, 288–298.

[44] Kim, K.H., Kim, J.H., Yoon, J., Y, J.K., 2008. Influence of task difficulty on the features of event-related potential during visual oddball task.

Neuroscience Letters 445, 179–83.

[45] King, J.R., Faugeras, F., Gramfort, A., Schurger, A., El Karoui, I., Sitt, J.D., Rohaut, B., Wacongne, C., Labyt, E., Bekinschtein, T., Cohen, L., Naccache, L., Dehaene, S., 2013. Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. NeuroImage 83, 726–738.

[46] Kiss, M., Van Velzen, J., Eimer, M., 2008. The N2pc component and its links to attention shift and spatially selective visual processing. Psychophysiology 45, 240–249.

[47] Kutas, M., McCarthy, G., Donchin, E., 1977. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197, 792–5.

[48] Labbé, B., Tian, X., Rakotomamonjy, A., 2010. MLSP competition, 2010: Description of the third place method. IEEE Int. Workshop on Machine Learning for Signal Processing , 116–117.

[49] Lance, B.J., Kerick, S.E., Ries, A.J., Oie, K.S., McDowell, K., 2012. Brain computer interface technologies in the coming decades, in: Proc. of the IEEE, pp. 1585–1599.

[50] Leiva, J.M., Martens, S.M.M., 2010. MLSP competition, 2010: Description of the first place method. IEEE Int. Workshop on Machine Learning for Signal Processing , 112–113.

[51] Luck, S.J., 2004. Ten simple rules for designing and interpreting ERP experiments, in: Handy, T.C. (Ed.), Event-Related Potentials: A Methods Handbook, pp. 1–8.

[52] Luck, S.J., 2005. A Introduction to the Event-Related Potential Technique. The MIT Press, Cambridge, MA.

[53] Luck, S.J., Hillyard, S.A., Mouloua, M., Woldorff, M.G., Clark, V.P., Hawkins, H.L., 1994. Effect of spatial cueing on luminance detectability: Psychophysical and electrophysiological evidence for early selection. J. of Experimental Psychology: Human Perception and Performance 20, 887–904.

[54] Luck, S.J., Kappenman, E.S., 2011. The Oxford Handbook of Event-Related Potential Components. Oxford University Press, USA.

[55] MacKay, D.J.C., 1992. Bayesian interpolation. Neural Comput. 4, 415–447.

[56] Makeig, S., Bell, A.J., Jung, T.P., Sejnowski, T.J., 1996. Independent component analysis of Electroencephalographic data, in: Advances in Neural Information Processing Systems 8, pp. 145–151.

[57] Marathe, A.R., Ries, A.J., Lawhern, V.J., Lance, B.J., Touryan, J., McDowell, K., Cecotti, H., 2015. The effect of target and non-target similarity on neural classification performance: a boost from confidence. Frontiers in Neuroscience 9, 1–11.

[58] Marathe, A.R., Ries, A.J., McDowell, K., 2014. Sliding HDCA: Single-trial eeg classification to overcome and quantify temporal variability. IEEE Trans. on Neural Systems and Rehab. Eng. 22, 201–211.

[59] Matran=Fernandez, A., Poli, R., 2016. Brain-computer interfaces for detection and localisation of targets in aerial images. IEEE trans. Biomed. Eng. , 1–10.

[60] Millán, J.d.R., Rupp, R., Müller-Putz, G.R., Murray-Smith, R., Giugliemma, C., Tangermann, M., Vidaurre, C., Cincotti, F., Kübler, A., Leeb, R., Neuper, C., Müller, K.R., Mattia, D., 2010. Combining braincomputer interfaces and assistive technologies: state-of-the-art and challenges. Frontiers in Neuroscience 4, 1–15.

[61] Müller, K.R., Anderson, A., Birch, G.E., 2003. Linear and nonlinear methods for brain-computer interfaces. IEEE Trans. Neural Syst. Rehabil. Eng. 11, 165–169.

[62] Mur, M., Bandettini, P.A., Kriegeskorte, N., 2009. Revealing representational content with pattern-information fMRI-an introductory guide. Soc. Cogn. Affect. Neurosci. 4, 101–109.

[63] Parra, L., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A., Sajda, P., 2003. Single trial detection in EEG and MEG: Keeping it linear. Neurocomputing 52-54, 177–183.

[64] Parra, L.C., Christoforou, C., Gerson, A.D., Dyrholm, M., Luo, A., Wagner, M., Philiastides, M.G., Sajda, P., 2008. Spatio-temporal linear decoding of brain state: Application to performance augmentation in high-throughput tasks. IEEE Signal Process. Mag. 25, 95–115.

[65] Pohlmeyer, E.A., Jangraw, D.C., Wang, J., Chang, S.F., Sajda, P., 2010. Combining computer and human vision into a BCI: can the whole be greater than the sum of its parts?, in: Proc. of the 32nd Int. IEEE EMBC Conf., Buenos Aires, Argentina. pp. 138–41.

[66] Pohlmeyer, E.A., Wang, J., Jangraw, D.C., Lou, B., Chang, S., Sajda, P., 2011. Closing the loop in cortically-coupled computer vision: a brain-computer interface for searching image databases. J. Neural Eng. 8, 036025.

[67] Polich, J., 2007. Updating P300: An integrative theory of P3a and P3b. Clinical Neurophysiology 118, 2128–2148.

[68] Potter, M.C., 1976. Short-term conceptual memory for pictures. J. of Experimental Psychology: Human Learning and Memory 2, 509–522.

[69] Prinzmetal, W., Zvinyatskovskiy, A., Gutierrez, P., Dilem, L., 2009. Voluntary and involuntary attention have different consequences: The effect of perceptual difficulty. The quaterly journal of experimental psychology 62, 352–369.

[70] Rakotomamonjy, A., Guigue, V., 2008. BCI competition iii : Dataset ii - ensemble of SVMs for BCI P300 speller. IEEE Trans. Biomed. Eng. 55, 1147–1154.

[71] Rivet, B., Souloumiac, A., 2013. Optimal linear spatial filters for event-related potentials based on a spatio-temporal model: Asymptotical performance analysis. Signal Processing 93, 387–398.

[72] Rivet, B., Souloumiac, A., Attina, V., Gibert, G., 2009. xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. IEEE Trans Biomed. Eng. 56, 2035–43.

[73] Sajda, P., Gerson, A., Parra, L., 2003. High-throughput image search via single-trial event detection in a rapid serial visual presentation task. In Proc. of the 1st Int. IEEE EMBS Conf. Neural Eng. , 7–10.

[74] Shannon, C.E., Weaver, W., 1964. The mathematical theory of communication. Urbana, IL: University of Illinois Press .

[75] Sirvent Blasco, J.L., Iáñez, E., Úbeda, A., Azorín, J.M., 2012. Visual evoked potential-based brain-machine interface applications to assist disabled people. Expert Systems with Applications 39, 7908–18.

[76] Smulders, F., Kenemans, J.L., Kok, A., 1994. A comparison of different methods for estimating single-trial P300 latencies. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section 92, 107–114.

[77] Sutton, S., Braren, M., Zubin, J., R., J.E., 1965. Evoked-potential correlates of stimulus uncertainty. Science 26, 1187–8.

[78] Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381, 520–2.

[79] Touryan, J., Gibson, L., Horne, H.J., Weber, P., 2011. Real-time measurement of face recognition in rapid serial visual representation. Frontiers in Psychology 2, 1–8.

[80] U.S Department of Defense Office of the Secretary of Defense, 1999. Code of federal regulations, protection of human subjects. Technical Report 32 CFR 219. Washington, DC: Government Printing Office.

[81] U.S. Department of the Army, 1990. Use of volunteers as subjects of research. Technical Report AR 70-25. Washington, DC: Government Printing Office.

[82] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M., 2002. Brain-computer interfaces for communication and control. Clin Neurophysiol 113, 767–791.

[83] Yu, K., Ai-Nashash, H., Thakor, N., Li, X., 2014. The analytic bilinear discrimination of single-trial eeg signals in rapid image triage. PLoS One 9, e100097.